

## Summary of ABRF samples analysis

Samples A and B were diluted in 50 uL and 100 uL of 50 mM Ammonium bicarbonate buffer, pH 7.8 with 0.1% pps silent surfactant (Protein Discovery), correspondently. For reduction and alkylation first DTT was added to a final concentration of 5 mM, and samples were incubated for 45 min at 50 C, then iodoacetamide was added to a final concentration of 15 mM and samples were incubated in the dark at room temperature for 60 min. For first digest trypsin (Promega) was added in 1: 22 enzyme: protein ratio and incubated for 4 hrs at 37C. then samples were concentrated in the Speed Vac: sample A up to 10 uL and sample B up to 20 uL. For second digest used 5 uL of each sample, diluted up to 80 uL with 50 mM Ammonium bicarbonate buffer, pH 7.8, added 100 ng of GluC (Roche Analytical) and incubated for 20 hrs at room temperature. Double digested samples were concentrated by Speed Vac up to 5 ul. Before mass spec analysis 5 uL of 1.25 M HCL was added to each sample for pps surfactant hydrolysis and samples were incubated for 1 hrs at room temperature. Samples were analyzed by nanoLC/MS/MS using 25 cm x 75 um Dionex PepMap C18 column, 2 D Exsigent LC pumps and LTQ Orbitrap XL. Bioworks™ 3.3.1 SP1 with SEQUEST® and Mascot™ software 2.2 (Matrix Sciences) were used for protein ID based on probability of 99.9 (p<0.01) and precursor mass accuracy of 5 ppm. Human IPI and UniProt databases were used for both search engines.

### Sample A

**Main Human recombinant protein** was identified either as truncated form of soluble form of receptor for advanced glycation endproducts [Precursor],

#### **Q86SN1\_HUMAN**

UniProt accession number **Q86SN1**

Q86SN1 full length is 347 AA, identified form is 23-327 plus GSHM on N-terminus.

Sequence coverage was 98.38% for truncated form and 92.35% for full protein

or as truncated form of advanced glycosylation end product-specific receptor precursor

#### **RAGE\_HUMAN**

UniProt accession number Q15109

Q15109 full length is 404 AA, identified form is 23-327 plus GSHM on N-terminus.

Sequence coverage was 98.38% for truncated form and 76.49% for full protein

**GSHMAQNITARI GEPLVLKCKG APKKPPQRL E**  
**WKLNTGRTEA WKVLSPOGGG PWDSVARVLP NGSFLPAVG IQDEGIFRCQ**  
**AMNRNGKETK SNYRVRVYQI PGKPEIVDSA SELTAGVPNK VGTCVSEGSY**

PAGTLSWHL D GKPLVPNEKG VSVKE<sub>qtr</sub>RH PETGLFTLQS ELMVTPARGG  
DPRPTFSCSF SPGLPR<sub>hr</sub>AL RTAPIQPRVW EPVPLEEVQL VVEPEGGAVA  
PGGTVTLTCE VPAQPSPQIH WMKDGVPLPL PPSVLILPE IGPQDQGTYS  
CVATHSSHGP QESRAVSISI IEPGEEG

*Other identified proteins were keratins (at least 4 different types) and ribosomal proteins from E.coli*

## Sample B

**Main Human recombinant protein** was identified either as truncated form of soluble form of receptor for advanced glycation endproducts [Precursor], **Q86SN1\_HUMAN** same as in **sample A**  
UniProt accession number **Q86SN1**

Q86SN1 full length is 347 AA, identified form is 23-327 plus GSHM on N-terminus.

Sequence coverage was 98.38% for truncated form and 92.35% for full protein

or as truncated form of advanced glycosylation end product-specific receptor precursor

**RAGE\_HUMAN** same as in **sample A**  
UniProt accession number Q15109

Q15109 full length is 404 AA, identified form is 23-327 plus GSHM on N-terminus.

Sequence coverage was 98.38% for truncated form and 76.49% for full protein

GSHMAQNITARI GEPLVLKCKG APKKPPQRL E  
WKLNTGRTEA WKVLS PQGGG PWDSVARVLP NGSFLFLPAVG IQDEGIFRCQ  
AMNRNGKETK SNYRVRVYQI PGKPEIVDSA SELTAGVPNK VGTCVSEGSY  
PAGTLSWHL D GKPLVPNEKG VSVKE<sub>qtr</sub>RH PETGLFTLQS ELMVTPARGG  
DPRPTFSCSF SPGLPR<sub>hr</sub>AL RTAPIQPRVW EPVPLEEVQL VVEPEGGAVA  
PGGTVTLTCE VPAQPSPQIH WMKDGVPLPL PPSVLILPE IGPQDQGTYS  
CVATHSSHGP QESRAVSISI IEPGEEG

**Another major protein identified only in Sample B was PRB4\_HUMAN**, Basic salivary proline-rich protein 4 allele M or S.

UniProt accession number **P10163**

Sequence coverage was 20% for short allele (247 AA full protein)

mlllllsva l lalssaesss edvsqeESLF LISGKPEGRr pgggnqarp  
ppppgKPQGP Ppgggngsqg pppppgkpeg rppggngsq gppphpgkpe  
rpppggngsq ggtppppgkp erpppggngq shrpppppgk perpppggn

qsqgppphpg kpegpppgeg nksrsarSPP GKPQGPPQQE GNKPQGPPPPp  
gkpqgpppag gnpqqpqpP AGKPQgpppp pgggrpprpa qgqppp

We also identified another proline rich protein 4 (PROL4\_HUMAN) as a minor component

UniProt accession number **Q16378**

Sequence coverage was 20% for full protein (134 AA)

mllvllsvl lalssaqstd ndvnyedftf tipdvedssq rpdQGPQRpp  
pegllprPPG DSGNQddgpq qrppkpgghh rhpppppfqn qqrpprrghr  
QLSLPRFPSV SLQEassffr rdrparhpqe qplw

*Other identified proteins were keratins (at least 4 different types) and ribosomal and heat shock proteins from E.coli same as in sample B*